

Japanese Patent, Laid-Open Publication No. S63-262693

Date of Publication: 28 Oct. 1988

Date of Application: 20 Apr. 1987 Application No: S62-97779

Applicant: Nippon Electric Co., Ltd.

Inventor: Toshio YOSHIKAWA

Patent Specification

1. Title of Invention

Speech discrimination and detection Apparatus

2. Claims

A speech discrimination and detection apparatus comprising:

a conversion circuit for converting an inputted signal contained in each of extraction periods of a pre-assigned length into line spectrum pair coefficients;

a coefficient distance determining circuit for determining whether the distance between a pair of the mutually adjacent line spectrum pair coefficients is longer than a pre-assigned threshold distance; and

a voiced sound discriminating circuit for discriminating a speech sound by judging whether the determination results of the coefficient distance determining circuit remained the same for more than a pre-determined length of time.

3. Detailed Description of the Invention

[Field of the industrial application of the Invention]

The invention is related with an apparatus for discriminating an inputted speech sound and in particular with a speech discrimination and detection apparatus incorporated into an apparatus such as a speech recognition apparatus and discriminating portions containing parts of a speech, in association with an inputted sound.

[Description of the Prior Art]

A speech determination and selection apparatus for discriminating and detecting speech portions, according to a prior art, is as shown in Fig.4 and comprises:

a circuit 8 for extracting from the inputted signal 1 which corresponds to an inputted sound, its level.

a threshold setting circuit 10 for pre-assigning the threshold value as a value determined in a manner related to parameters such as the noise level and/or the speech level at a time of the sound being inputted to and issuing determination signals 11 notifying

starting and ending points of a speech portion by determining a relevant point as the starting point of a speech portion by comparing the inputted level signal 9 made available by the circuit 8 with a threshold value and when measuring the inputted level signal 9 to stay continuously higher than the threshold value for more than a pre-assigned time, and determining a relevant point as the end point of the speech portion by comparing the inputted level signal 9 with the threshold value and when measuring the inputted level signal 9 to stay continuously lower than the threshold value for more than another pre-assigned time, and

a speech portion detecting circuit 12 for issuing speech detection result indicating signals 7 in response to a receipt of the determination signals 11 notifying starting and ending points of the speech portion.

Speech portions of an input signal are determined as portions each of which is defined as a period between a pair of starting and ending points determined as above according to a prior art technology.

[Problems to be solved by the Invention]

The above explained prior art speech discrimination and detection apparatus relies on information associated with the power level of an inputted sound and as a result of this it is prone to collecting environmental noise and associated with a problem, difficulty in separating between the inputted sound and the environmental noise..

[Means for solving the problem]

A speech discrimination and detection apparatus of the present invention comprises:

a conversion circuit for converting an inputted signal contained in each of extraction periods of a pre-assigned length into line spectrum pair coefficients;

a coefficient distance determining circuit for determining whether the distance between a pair of the mutually adjacent line spectrum pair coefficients is longer than a pre-assigned threshold distance; and

a voiced sound discriminating circuit for discriminating a speech sound by judging whether the determination results of the coefficient distance determining circuit remained the same for more than a pre-determined length of time.

[Operation]

According to the present invention, therefore, it becomes possible to discriminate speech portions from an input signal, even under presence of environmental noise, because the present invention is configured so that distances between line spectrum pair

coefficients are analyzed for determining speech portions of the input signal.

[Embodiment of the Invention]

Below is an explanation about an embodiment of the present invention and is given along related drawings.

Fig.1 is a block diagram associated with an embodiment of the present invention and Figs.2 and 3 respectively are charts for explaining the embodiment in Fig.1. An inputted signal 1 corresponding to an inputted sound is expected in general to contain environmental noise. The line spectrum pair converting circuit 2 converts an inputted signal 1 into a line spectrum pair (herein after called LSP) coefficient signal 3 in accordance to the LSP method which belongs to a kind of the linear prediction coding method. Here, the LSP coefficients are parameters of a frequency domain.

For instance, calculating in 8th-order of analysis order, 8 parameters, $w_1, w_2, w_3, \dots, w_8$ are obtained as shown in Figs.2 and 3. Here, the analysis is made for the sampling frequency of 8kHz, at the frequency band of 0.4 – 3.4 kHz, the same frequency band as that of the telephone, and for the analysis frame period between 10 – 20 msec.

Further detail on the LSP may be found in an article titled "Speech synthesis method using line spectrum frequencies as parameters and speech synthesis LSI", Nikkei Electronics, p.p. 128 – 158 Issue No. 257, 02 Feb. 1981.

LSP coefficients, $w_1 - w_p$ are parameters of a frequency domain and they tend to gather in the vicinities of formant frequencies $F_1 - F_{P/2}$ associated with a voiced sound. Here, the following relationship holds among the LSP coefficients.

$$0 < w_1 < w_2 < \dots < w_{p-1} < w_p < \square$$

with P being the analysis order.

The coefficient distance determining circuit 4 calculates distances between adjacent line spectrum coefficient pairs, $(w_2 - w_1) \dots (w_7 - w_8)$, based on a line spectrum pair coefficient signal 3, as shown in Fig.2.

Below is an example of calculation methods for the coefficient distances, $(w_n - w_{n-1})$.

Assuming the LSP analysis is performed in a P-th order of the analysis order, following formulas are calculated for cases in which $n = 2, 3, \dots, P$, respectively.

$$(w_n - w_{n-1}) < w_{TH1} \quad \dots \quad (1)$$

$$(w_n - w_{n-1}) < w_{TH2} \quad \dots \quad (2)$$

Here, w_{TH1} and w_{TH2} are threshold values associated with the LSP coefficient distances $(w_n - w_{n-1})$, respectively and hold the relation, $w_{TH1} < w_{TH2}$.

When determined that there is one or more LSP coefficient w_1, \dots, w_p satisfying formula (1) and in addition, there are two or more LSP coefficients w_1, \dots, w_p satisfying

formula (2), the coefficient distance determination result signal 5 is judged to be of a voiced sound and the voiced sound determining circuit 6 issues a speech determination result signal 7 after receiving, for 3 consecutive frames, the coefficient distance determination result signal 5 of indicating the receipt of a voiced sound.

Fig.2 shows the relation between the frequency spectrums and the LSP coefficients corresponding to a situation in which a voiced sound is inputted in the embodiment case shown in Fig.1. Fig.3 shows the relation between the frequency spectrums and the LSP coefficients corresponding to a situation in which an unvoiced sound or an environmental noise is inputted in the same embodiment case shown in Fig.1.

As it becomes clear from Fig.2, when the inputted sound is a voiced sound, the LSP coefficients w_1, \dots, w_8 are positioned in the vicinities of formant frequencies, F_1, \dots, F_4 . In particular, for the first formant frequency, F_1 , as the resonance gain is larger in general, LSP coefficients w_1 and w_2 are positioned more closely to the formant frequency, and hence, the LSP coefficient distance ($w_2 - w_1$) becomes shorter than the threshold w_{TH1} and the LSP coefficient distance ($w_4 - w_3$) near the second formant frequency becomes shorter than the threshold w_{TH2} .

In contrast, when the inputted sound is an unvoiced sound or an environmental noise, the associated frequency spectrums are very flat as shown in Fig.3 and the LSP coefficients w_1, \dots, w_8 are not positioned very close to those spectrums. As a result of this, the LSP coefficient distances, ($w_n - w_{n-1}$) do not become shorter than either of w_{TH1} and w_{TH2} .

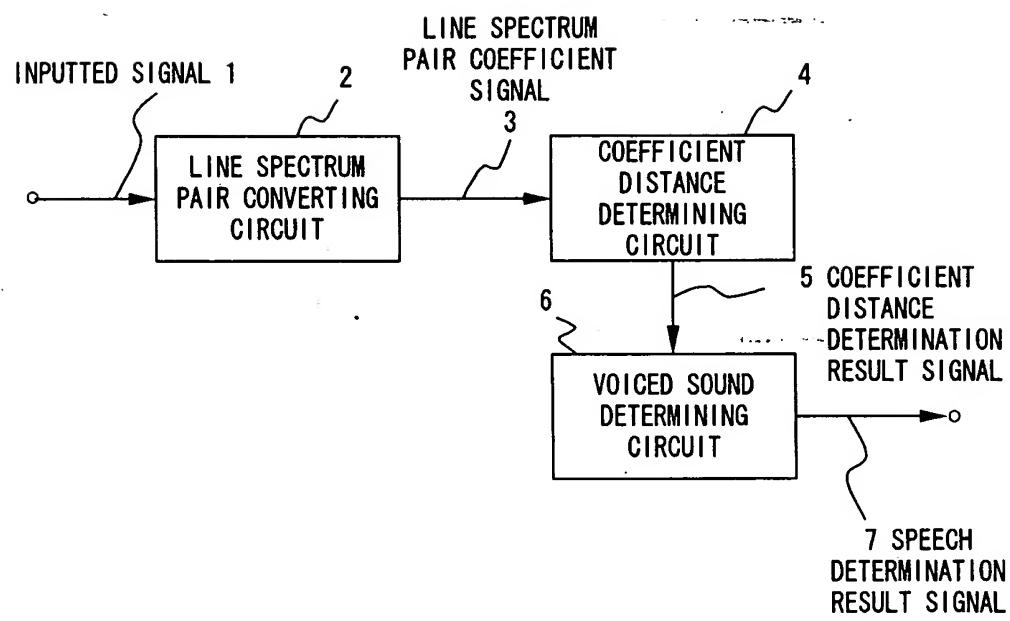
[Effect of the invention]

As described above, the present invention is concerned with a configuration in which an inputted signal is determined whether it is of a speech signal by a method of analyzing line spectrum pair distances in place of relying on the method, comparing the levels of an inputted signal so that it becomes possible to extract a speech signal from an inputted signal even if the inputted signal is composed of a mixture of speech and environmental noise sounds, provided that the inputted signal is a voiced sound. Incorporating the invention apparatus into a speech recognition apparatus is effective in improving its speech recognition success rate.

4. Brief explanation of drawings

Fig.1 is a block diagram showing an embodiment of the present invention. Fig.2 and Fig.3 are charts associated with the embodiment shown in Fig.1. Fig.4 is a block diagram showing an embodiment of a prior art technology.

2: line spectrum pair converting circuit 4: coefficient distance determining circuit
6: voiced sound determining circuit



F I G. 1

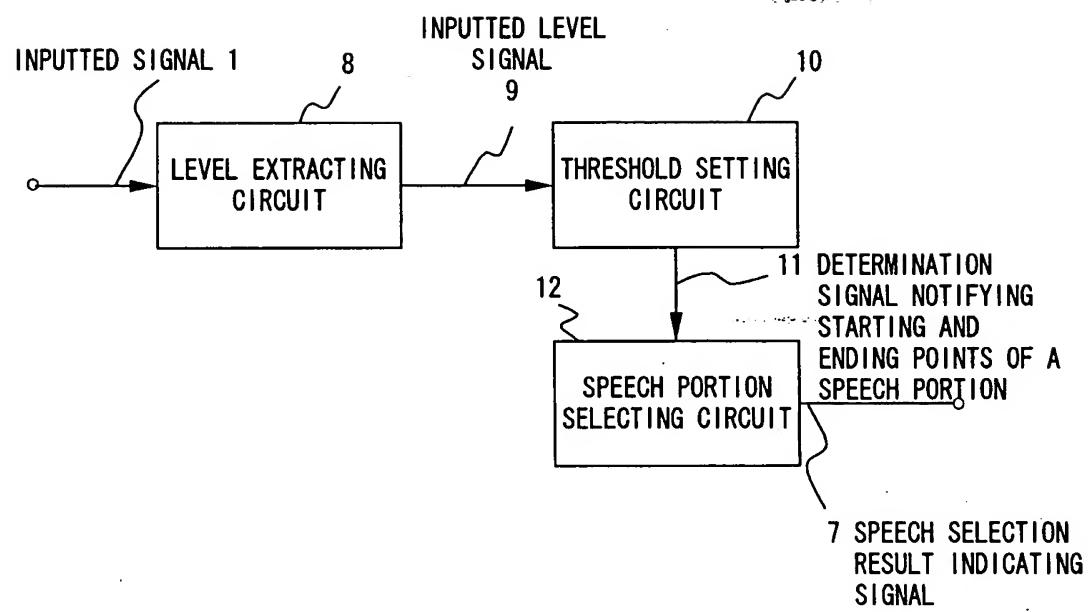


FIG. 4

⑩ 日本国特許庁 (JP) ⑪ 特許出願公開
⑫ 公開特許公報 (A) 昭63-262693

⑬ Int.Cl. 4 識別記号 庁内整理番号 ⑭ 公開 昭和63年(1988)10月28日
G 10 L 3/00 9/14 3 0 1 A-7627-5D
D-8622-5D
3 0 1 A-7627-5D 審査請求 未請求 発明の数 1 (全4頁)

⑮ 発明の名称 音声判定検出装置

⑯ 特願 昭62-97779

⑰ 出願 昭62(1987)4月20日

⑱ 発明者 吉川 敏雄 東京都港区芝5丁目33番1号 日本電気株式会社内

⑲ 出願人 日本電気株式会社 東京都港区芝5丁目33番1号

⑳ 代理人 弁理士 内原 晋

明細書

1. 発明の名称

音声判定検出装置

2. 特許請求の範囲

入力信号を一定の抽出区間ごとに級スペクトル対係数に変換する変換回路と、

該変換された級スペクトル対係数の隣接する係数間の距離が、しきい値より大きいか小さいかを判定する係数間距離判定回路と、

該係数間距離判定回路の判定結果が、連続して一定時間以上維持したかどうかを判定し音声を検出する有音声判定回路と、を備えている音声判定検出装置。

3. 発明の詳細な説明

〔産業上の利用分野〕

本発明は入力する音声を判定して検出する装置に関し、とくに音声認識装置などにおける入力音声の存在範囲を判定し検出する音声判定検出装置に関する。」

〔従来の技術〕

従来、音声区間を判定検出する音声判定検出装置は、第4図に示されるように、入力する音に対応する入力信号1からレベルを抽出する回路8と、音声入力時の雜音レベル、ならびに入力音声レベルなどによってレベルのしきい値を設定し、該しきい値と、前記レベル抽出回路8から送出される入力レベル信号9とを比較して、該入力レベル信号9が大である状態が、定められた一定時間以上維持したとき、音声区間の始端と判定し、その後、前記しきい値と当該入力レベル信号9とを比較して、該入力レベル信号9が小である状態が、定められた一定時間以上維持したときに、音声区間の終端と判定して、音声区間の始端ならびに終端の判定信号11を送出するしきい値設定回路10と、該音声区間始端終端判定信号11を入力して、音声検出の結果信号7を送出する音声区間の検出回路12と、を備えていて、上述の判定により決定された始端から終端までを、音声区間として検出していた。

〔発明が解決しようとする問題点〕

上述した従来の音声判定検出装置は、入力音声のパワー情報を用いるため、周囲雜音が混入しやすく、入力音声と周囲雜音との区別が困難という欠点がある。

〔周囲点を解決するための手段〕

本発明の音声判定検出装置は、入力信号を一定の抽出区間ごとに線スペクトル対係数に変換する変換回路と、該変換された線スペクトル対係数の隣接する係数間の距離が、しきい値より大きいか小さいかを判定する係数間距離の判定回路と、該係数間距離判定回路の判定結果が、連続して一定時間以上続続したかどうかを判定し音声を検出する有聲音判定回路と、を備えている。

〔作用〕

したがって本発明によると、入力信号が音声信号であるか否かの判定に、線スペクトル対係数の係数間距離を用いるため、周囲雜音があっても音声を判定し検出することができる。

〔実施例〕

以下に本発明を、その実施例について図面を参考して説明する。

3

る。

LSP係数 $w_1 \sim w_p$ は、周波数領域のパラメータであって、音声のホルマント周波数 $F_1 \sim F_{p/2}$ の近傍に集中するという性質があり、また、各LSP係数 $w_1 \sim w_p$ には、次の関係が成立している。すなわち、 $0 < w_1 < w_2 < \dots < w_{p-1} < w_p < \pi$ であり、ここで p は分析次数である。

この性質を利用して、係数間距離の判定回路4により、線スペクトル対係数信号3にて、第2図のように隣接するLSP係数間の距離 ($w_2 - w_1$) \sim ($w_p - w_{p-1}$) を計算する。

係数間距離 ($w_n - w_{n-1}$) の計算方法の一例をつぎに述べる。LSP分析の次数が P 回のとき、 $n = 2, 3, \dots, p$ において、次式を計算する。

$$(w_n - w_{n-1}) < w_{th1} \quad \cdots \cdots (1)$$

$$(w_n - w_{n-1}) < w_{th2} \quad \cdots \cdots (2)$$

なお、 w_{th1} と w_{th2} とは、LSP係数間距離 ($w_n - w_{n-1}$) のしきい値であり、 $w_{th1} <$

照して説明する。

第1図は本発明による一実施例を示すブロック図、第2図ならびに第3図はそれぞれ、同上の実施例を説明するグラフ図である。入力する音に対応する入力信号1は、通常、周囲雜音を含んでいる。線スペクトル対の変換回路2は、入力信号1を、線形予測符号化法の一様である線スペクトル対 (Line Spectral Pair、以下、LSPと称す) 方式により、周波数領域のパラメータである線スペクトル対 (LSP) 係数の信号3に変換する。

たとえば、LSP係数は、分析次数を8次で計算すると、第2図、第3図の如く、 $w_1, w_2, w_3, \dots, w_p$ の8個が求められる。なお、分析は、標準化周波数が8kHzで、帯域幅を電話帯域の0.4~3.4kHzとし、分析フレーム周期を10~20msとする。

また、LSPについては、1981年2月2日発行の「日経エレクトロニクス」No.257の記事「線スペクトル周波数をパラメータとした音声合成法とそのLSP化」P.P. 128~158に解説されている。

4

w_{th2} に設定される。

(1) 式を満足するLSP係数 $w_1 \sim w_p$ が1個以上存在し、かつ (2) 式を満足するLSP係数 $w_1 \sim w_p$ が2個以上存在すれば、係数間距離判定結果の信号5が、有聲音であると判定され、次に有聲音の判定回路6は、係数間距離判定結果信号5が有聲音であることを、たとえば連続して3フレーム継続して入力されると、音声検出結果の信号7を出力する。

第2図は、第1図の実施例において、有聲音の場合の周波数スペクトルとLSP係数との関係を示し、また第3図は、第1図の実施例において、無聲音あるいは周囲雜音の周波数スペクトルとLSP係数との関係を示す。

第2図から分かるように、有聲音の場合、ホルマント周波数 $F_1 \sim F_{p/2}$ の近傍にLSP係数 $w_1 \sim w_p$ が集中している。また、第1ホルマント周波数 F_1 は一般に共振の利得が高いため、LSP係数 w_1, w_2 の集中度も強まって、LSP係数間距離 ($w_2 - w_1$) は、しきい値 w_{th1}

5

-706-

6

より小さくなり、第2ホルマント周波数 w_2 近傍のLSP係数間距離 ($w_4 - w_3$) は大きい値 w_{TM2} よりも小さくなる。

しかし無聲音や周囲雜音の場合、第3図の如く、周波数スペクトルが平坦であり、LSP係数 $w_1 \sim w_8$ の集中は少ない。このため、LSP係数間距離 ($w_{TM1} \sim w_{TM8}$) は大きい値 $w_{TM1} \sim w_{TM2}$ より小さくなることはない。

【発明の効果】

以上説明したように本発明は、入力信号が音声信号であるかどうかを判定するために、入力信号レベルの大きさで判定するかわりに、級々ペクトル係数の係数間距離を用いることにより、周囲雜音にうまれた音声でも、有聲音であれば検出することが可能であるから、音声認識装置における認識率の向上に効果がある。

4. 図面の簡単な説明

第1図は本発明による一実施例を示すブロック図、第2図ならびに第3図は、それと同上を説明するためのグラフ図、第4図は従来例を示すブ

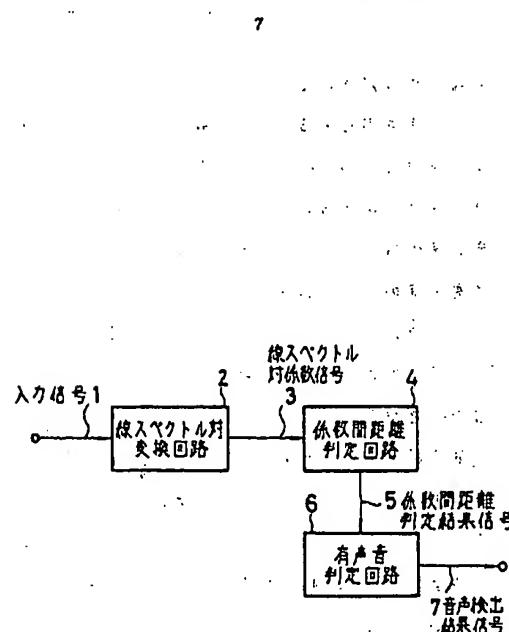
ロック図である。

2…級スペクトル対変換回路、

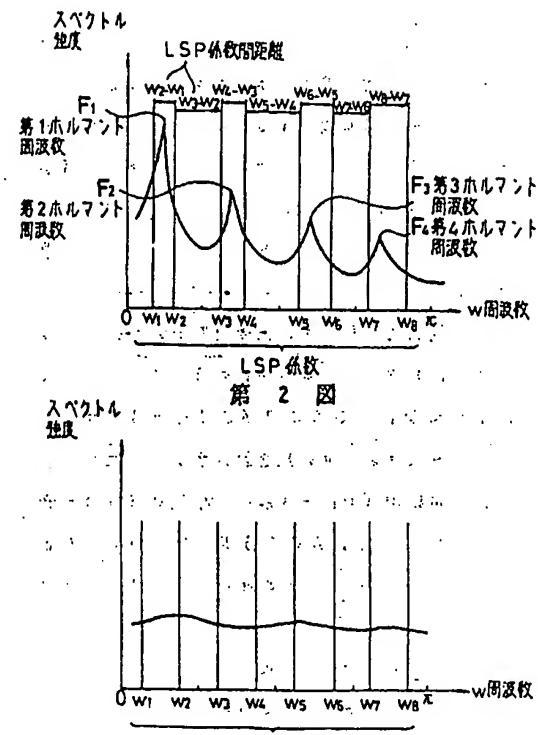
4…係数間距離判定回路、

6…有聲音判定回路。

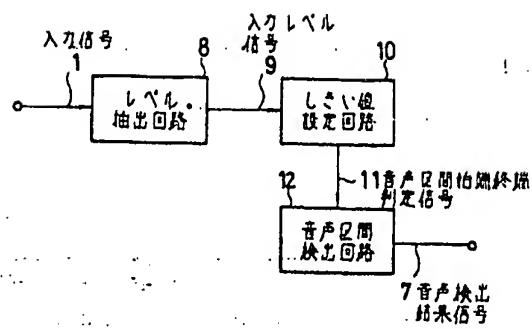
特許出願人 日本電気株式会社
代理人弁理士 内原 勝



第1図



第3図



第 4 図